ELSEVIER Theoretical Computer Science 290 (2003) 2137–2146

Theoretical Computer Science

www.elsevier.com/locate/tcs

Note

Weak repetitions in Sturmian strings

Ayşe Karaman

School of Computing, Queen's University, Kingston, Ont., Canada K7L 3N6

Received 27 August 2001; received in revised form 4 June 2002; accepted 18 June 2002
Communicated by G. Rozenberg

Abstract

In this paper, we analyze the weak repetitions in Sturmian strings and show that an optimally efficient algorithm to compute the weak repetitions in Sturmian strings according to the output encodings defined in the literature is quadratic in the string length. Finally, we present an encoding that leads to a linear-time algorithm.

Keywords: Sturmian strings; Weak repetitions; Encodings

1. Introduction

Sturmian strings are the family of right infinite strings on a binary alphabet with the appealing property that they have the minimum subword complexity and are not ultimately periodic. The representations and combinatorial properties of Sturmian strings have been studied by various authors [1–4,7–9,11].

A string is said to be a weak repetition if it is a concatenation of two or more finite strings each of which is a permutation of the other. An optimal algorithm to compute the weak repetitions in an arbitrary string to produce an output according to the encodings defined in the literature is quadratic in the string length [6].

In this paper, we analyze the weak repetitions in Sturmian strings based solely on their balance property. Our results indicate that the number of weak squares in a finite Sturmian string is quadratic in the string length. We also investigate the representation of weak repetitions according to the output encodings defined in the literature, namely C-encoding [5] and R-encoding [6], and show that the optimal efficiency of an algorithm to compute the weak repetitions in a Sturmian string using anyone of these

E-mail address: karaman@cs.queensu.ca (A. Karaman).

PII: S0304-3975(02)00574-1

encodings is still quadratic in the string length. Thus, the complexity of an algorithm to solve this problem does not improve when the input domain is restricted to Sturmian strings. Finally, we present an output encoding that leads to the linear time computation of weak repetitions in this family of strings. Following the review of terminology and notation in the next section, we present our analysis of weak squares in Section 3. Section 4 includes our results related to the output encodings existing in the literature. In the last section, we present the output encoding we are proposing along with the concluding remarks. For a detailed description of our findings, also see Karaman [10].

2. Preliminaries

A *string* is a sequence of letters from a particular set called *alphabet*. Throughout this paper, we will be studying on a *binary* alphabet $A=\{a,b\}$, an alphabet with 2 letters. A string is either finite, i.e., a sequence of finite number of letters, or infinite at the right end. We use the notation A^* and A^ω, respectively to denote the finite and infinite set of strings on alphabet A. The *empty string*, ε, is the string with no letters. We represent by $x[i]$ the ith letter of the string x. The *length* of a finite string x, denoted by $|x|$, is the number of letters in x. Note that $|\varepsilon|=0$. Given the strings $x\in A^*$ and $y\in A^*\cup A^\omega$, xy is the concatenation of x and y. If a string y is of the form $y=uxv$, then x is called a *substring* of y. Particularly, in this case, u is called a *prefix* and v is a *suffix* of y. ε is the substring of any string by definition. The notation $y[i..j]$ where $i,j\in[1,|y|]$ denotes the substring x of y so that $x[1]=y[i]$ and $|x|=j-i+1$. We represent by $|x|_l$ where $l\in A$ and $x\in A^*$ the number of occurrences of letter l in x. For a string $x\in\{a,b\}^*$, we call $|x|_b$ the *weight* of x. A binary string x is named as *ultimately periodic* if it is of the form $x=uv^\omega$ for any $u\in A^*$, $v\in A^*\backslash\{\varepsilon\}$, i.e., if it has a suffix that repeats itself infinitely many times.

Consider the *complexity function* [2] $P(x,i)$ which gives the number of distinct substrings of length i of the string x. For x non-ultimately periodic, we have $P(x,i)\geqslant i+1$ for all $i>0$ [4]. A *Sturmian string* s is a string in $\{a,b\}^\omega$ where $P(s,i)=i+1$, the minimum possible value for all $i>0$. Another way to identify the Sturmian strings is based on their *balance* property: a string x on a binary alphabet is defined as balanced if $||u|_b-|v|_b|\leqslant 1$ for all finite substring pairs u and v of x where $|u|=|v|$. An infinite balanced string is either Sturmian or ultimately periodic [4]. Then, a Sturmian string is one that is balanced and not ultimately periodic. A *finite Sturmian string* is defined as a finite substring of a Sturmian string. Throughout this paper, we assume without loss of generality that in a Sturmian string $s\in\{a,b\}$, letter a is the repeating letter.

A string x is said to be a *weak repetition* of *order* n if it is of the form $x=u_1u_2\dots u_n$, a concatenation of the substrings u_i for $i=1..n$ where

$$|u_i|_l=|u_j|_l \quad \forall\ l\in A,\ \ i,j=1..n.$$

In such a repetition, we call $|u_i|$ the *period length*, and each u_i the ith *occurrence of the period*. When a weak repetition $x=u_1\dots u_n$ is a substring of a string y so that $y=y_1xy_2$, the position m in y where $y[m]=x[1]=u_1[1]$, namely the position where the weak repetition starts, is the *position* of the repetition x in y. When a weak repetition is of order 2, we call it a *weak square*.

3. Analysis of weak repetitions

In this section we analyze the weak repetitions in Sturmian strings using solely their balance property. We first introduce a weak square table structure based on the results of the following lemma to encode the weak squares in a given Sturmian string. Then, we continue our analysis on this structure to count the number of weak squares in a Sturmian string. Before going further, we define a unary relationship ω on string $x \in A^*$ so that $\omega(x)=1$ if x is a weak square, $\omega(x)=0$ otherwise.

Lemma 1. *Let $s=uv$ be a finite Sturmian string where $|s|$ and $|u|$ are even. Then, $\omega(s)=1 \Leftrightarrow |s|_b$ is even. Furthermore, $\omega(u)=\omega(uv) \Leftrightarrow \omega(v)=1$.*

Proof. It is clear that s has to be of even length in order to be a weak square. From the balance property of Sturmian strings, the weights of first and second halves of a finite Sturmian string s can differ by either 0 or 1, and this difference is 0 if and only if $|s|_b$ is even, which is the case of a weak square. Looking into the substrings of even length of s, the weights of u and uv are both even or odd if and only if $|v|_b$ is even. Then $\omega(u)=\omega(uv) \Leftrightarrow \omega(v)=1$. □

Lemma 1 suggests that a Sturmian string can be examined, starting from the leftmost smallest possible weak square, namely the first two letters distinctively at even and odd positions, by iterating to check the next two letters to see whether there exists a letter b to invert the weak square condition of the current word. With this in mind, consider the definition of two-dimensional weak square tables $W_o(s)$ and $W_e(s)$ on a Sturmian string s as follows:

$$W_o(s)[1..\lfloor |s|/2 \rfloor, 1..\lfloor |s|/2 \rfloor]$$
$$W_o(s)[i,j] = \begin{cases} \omega(s[2i-1..2j]) & \text{if } i \leqslant j; \\ \text{nil} & \text{if } i > j. \end{cases}$$
$$W_e(s)[1..\lfloor (|s|-1)/2 \rfloor, 1..\lfloor (|s|-1)/2 \rfloor]$$
$$W_e(s)[i,j] = \begin{cases} \omega(s[2i..2j+1]) & \text{if } i \leqslant j; \\ \text{nil} & \text{if } i > j. \end{cases}$$

If we look at $W_o(s)$ table, row i encodes all weak squares positioned at $2i-1$, the table in turn encodes the entire weak squares at odd positions of the Sturmian string s. Similarly, $W_e(s)$ encodes the entire weak squares at even positions of s. Thus, $W_o(s)$ and $W_e(s)$ tables together give us all weak squares in s.

Throughout the remainder of this paper, we will be referring by W to a W_o or W_e table associated to a particular string since W_o and W_e have the same properties as one another in our context.

Lemma 2. *Let W be a weak square table of a Sturmian string s. Let $W[i,k]$ and $W[j,k]$ be two entries of W where $i \neq j$, $i \leqslant k$ and $j \leqslant k$. Then, $W[i,k+1] = W[j,k+1]$ if and only if $W[i,k] = W[j,k]$.*

Proof. The result of Lemma 1 implies that $\omega(u) = \omega(v) \Leftrightarrow \omega(uv) = 1$ for all u and v where uv is a finite Sturmian string with $|u|$ and $|v|$ even. Suppose $W = W_{\rm o}(s)$. Then, $W[i,k] = \omega(s[2k-1..2k]) \Leftrightarrow W[i,k+1] = 1$ and $W[j,k] = \omega(s[2k-1..2k]) \Leftrightarrow W[j,k+1] = 1$. This concludes that $W[i,k] = W[j,k] \Leftrightarrow W[i,k+1] = W[j,k+1]$. A similar proof applies for the case $W = W_{\rm e}$. $\square$

Suppose two distinct rows, i and j of a weak square table W. From Lemma 2, we immediately conclude that $W[i,m] = W[j,m]$ if and only if $W[i,n] = W[j,n]$ for any distinct columns m and n of W table where $m,n \geqslant max\{i,j\}$. In this case, we say that rows i and j follow an *identical pattern* if $W[i,m] = W[j,m]$, a *reverse pattern* otherwise, i.e., if $W[i,m] \neq W[j,m]$. Observe that, the sets of rows which follow distinctively the identical and reverse patterns as row 1 partition the entire set of rows of a W table into two. Based on this result, the $W_{\rm o}$ and $W_{\rm e}$ tables can be further encoded respectively into tables $W'_{\rm o}$ and $W'_{\rm e}$ as follows:

$W'_{\rm o}(s)[1..2, 1..\lfloor |s|/2 \rfloor]$

$$W'_{\rm o}(s)[i,j] = \begin{cases} W_{\rm o}(s)[i,j] & \text{if } i = 1, \\ 1 & \text{if } i = 2,\ j > 1 \text{ and row } j \text{ of } W_{\rm o}(s) \text{ follows} \\ & \qquad \text{an identical pattern as row 1;} \\ 0 & \text{if } i = 2,\ j > 1 \text{ and row } j \text{ of } W_{\rm o}(s) \text{ follows} \\ & \qquad \text{a reverse pattern as row 1;} \\ \text{nil} & \text{if } i = 2,\ j = 1. \end{cases}$$

$W'_{\rm e}(s)[1..2, 1..\lfloor (|s|-1)/2 \rfloor]$

$$W'_{\rm e}(s)[i,j] = \begin{cases} W_{\rm e}(s)[i,j] & \text{if } i = 1; \\ 1 & \text{if } i = 2,\ j > 1 \text{ and row } j \text{ of } W_{\rm e}(s) \text{ follows} \\ & \qquad \text{an identical pattern as row 1;} \\ 0 & \text{if } i = 2,\ j > 1 \text{ and row } j \text{ of } W_{\rm e}(s) \text{ follows} \\ & \qquad \text{a reverse pattern as row 1;} \\ \text{nil} & \text{if } i = 2,\ j = 1. \end{cases}$$

From these definitions, any two non-nil entries $W'(s)[2,i]$ and $W'(s)[2,j]$ of an encoded weak square table of a Sturmian string s are equal if and only if the rows i and j in the associated weak square table have the same pattern as one another.

Lemma 3. *Let $W = W(s)$ be a weak square table of a Sturmian string s. Then $W[i,i] = 1 \Leftrightarrow W[i,i+1] = W[i+1,i+1]$ $\forall$ $i \geqslant 1$.*

Proof. Consider with no loss in generality that $W = W_{\rm o}$. From the $W_{\rm o}$ table definition, we are looking at the following substrings $s[2i-1..2i]$, $s[2i-1..2i+2]$ and $s[2i+1..2i+2]$ of s of which the weak square conditions are encoded respectively into the entries $W[i,i]$, $W[i,i+1]$, and $W[i+1,i+1]$. Considering all possible cases leads us to the proof. $\square$

Lemma 4. *Let $W'(s)$ be the encoded weak square table of a Sturmian string s. Then, $W'(s)[2,i] = W'(s)[1,i-1]$ for all $i \geqslant 2$.*

Proof. Suppose without loss of generality that $W'(s) = W'_o(s)$. Consider the entry $W'(s)[1,i-1]$. There are two possibilities:

Case (a): $W'(s)[1,i-1] = 1$.

Case (b): $W'(s)[1,i-1] = 0$.

Case (a): $W'(s)[1,i-1] = 1$. In this case, if $W'(s)[1,i] = 1$, then since $W'(s)[1,i-1] = W'(s)[1,i] = 1$ and by Lemma 1, $|s[2i-1..2i]|_b = 0$. Thus, $W(s)[i,i] = W(s)[1,i] = 1$. This means that row i follows an identical pattern as row 1 in $W(s)$ and $W'(s)[2,i] = W'(s)[1,i-1] = 1$.

Suppose now that $W'(s)[1,i] = 0$. Then, $W'(s)[1,i-1] \neq W'(s)[1,i]$ and Lemma 1 imply that $|s[2i-1..2i]|_b = 1$. Thus, $W(s)[i,i] = W(s)[1,i] = 0$, meaning that row i follows an identical pattern as row 1 in $W(s)$ and $W'(s)[2,i] = W'(s)[1,i-1] = 1$.

Case (b): $W'(s)[1,i-1] = 0$. Following a similar argument as in the previous case, we conclude that $W(s)[i,i] \neq W(s)[1,i]$ therefore row i has always a reverse pattern as row 1 in $W(s)$, justifying the proof. □

According to Lemma 4, $W'[1,1..n] = W'[2,2..n+1]$ for all n where W' is an encoded weak square table. We use this result in Lemma 5 to show that the consecutive rows of a weak square table following an identical pattern as one another differ by exactly 1 in the number of weak squares encoded in each. To aid us in the remaining of this section, consider the notation $\Gamma_i(W)$ to denote the sequence of non-nil entries of row i of a given weak square table W.

Lemma 5. *Let $W = W(s)$ be a weak square table of a finite Sturmian string s and W' be the encoded table of W. Consider a row i in W. If there exists a row j in W where $i < j$, $W'[2,i] = W'[2,j]$, and $W'[2,k] \neq W'[2,i]$ for all $k : i < k < j$, then $|\Gamma_j(W)|_1 = |\Gamma_i(W)|_1 - 1$.*

Proof. The sequence $\Gamma_i(W)$ can be exactly one of the following:

(a) $1x$ where $x \in \{0,1\}^*$,
(b) 00^m1x where $m \geqslant 0$, 0^m is a sequence of m 0's, and $x \in \{0,1\}^*$,
(c) 00^m where $m \geqslant 0$, 0^m is a sequence of m 0's.

Case (a): $1x$ where $x \in \{0,1\}^*$. Then by Lemma 3, $W'[2,i] = W'[2,i+1]$. So $j = i+1$, $\Gamma_j(W) = x$ and $|\Gamma_j(W)|_1 = |\Gamma_i(W)|_1 - 1$.

Case (b): 00^m1x where $m \geqslant 0$, $x \in \{0,1\}^*$. Consider without loss of generality that $i = 1$. Then by Lemma 4, $W'[2,m+3] = 1$ and $W'[2,k] = 0$ for all $k < m+3$. Therefore $j = i+m+2$, $\Gamma_j(W) = x$ and $|\Gamma_j(W)|_1 = |\Gamma_i(W)|_1 - 1$.

Case (c): 00^m where $m \geqslant 0$. Consider again without loss of generality that $i = 1$. Then, by Lemma 4, there does not exist a row j in W satisfying the condition. □

Consider $\{i_1, i_2, \dots i_n\}$ where $i_1 < i_2 < \cdots < i_n$ as the set of entire rows that follow an identical pattern with one another in $W(s)$ where $W(s)$ is a weak square table of a finite Sturmian string s. Lemma 5 tells that $|\Gamma_{i_j}(W(s))|_1 = |\Gamma_{i_{j+1}}(W(s))|_1 + 1\ \forall j = 1..n-1$. From the definition of a weak square table, $|\Gamma_{i_n}(W(s))|_1$ is either 0 or 1, therefore, the total number of weak squares encoded in rows $i_1, i_2, \dots i_n$ is $\sum_{j=1}^{m} j$ where $m = |\Gamma_{i_1}(W(s))|_1$.

Lemma 6. *Let $W = W(s)$ be a weak square table of a Sturmian string s and W' be the encoded table of W. Consider the rows 1 and $i > 1$ in W where $W'[2,i] = 0$ and $W'[2,k] = 1$ for all $i > k > 1$. Then, $|\Gamma_i(W)|_1 = |\Gamma_1(W)|_0 - 1$.*

Proof. Since the rows 1 and i follow the reverse patterns as one another by condition, the non-nil entries of row i of W are 1 where the entries in the corresponding columns in row 1 are 0. Therefore, $|\Gamma_i(W)|_1 = |\Gamma_1(W)|_0 - n$ where n is the number of 0 entries in columns $1..i-1$ of row 1 of W. By condition, $W'[2,i] = 0$ and $W'[2,k] = 1$ for all $i > k > 1$. From this, Lemma 4 implies that $W'[1,k] = 1$ for all $i-1 > k > 1$ and $W'[1,i-1] = 0$. Since $W'[1,k] = W[1,k]\ \forall k$ by definition, $n = 1$, justifying the proof. □

Let $i > 1$ be the minimum index of rows which are following the reverse pattern as does row 1 in a weak square table $W(s)$ of a Sturmian string s. According to Lemma 6, the number of weak squares encoded in row i is one less than the number of 0 entries in row 1. This indicates the relationship between two rows following a reverse pattern with one another in a weak square table as of the number of weak squares encoded in each. With these findings, we are now ready for the following proof:

Theorem 1. *Let $W = W(s)$ be a weak square table of a finite Sturmian string s. Let $n = |\Gamma_1(W)|$ and $m = |\Gamma_1(W)|_1$. Then, the number of weak squares in s encoded in W is $(m(m+1) + (n-m)(n-m-1))/2$.*

Proof. From Lemma 5, the number of 1 entries encoded in rows that have an identical pattern as row 1 of W is

$$\sum_{i=1}^{m} i = m(m+1)/2.$$

And by Lemma 6, $|\Gamma_k(W)|_1 = |\Gamma_1(W)|_0 - 1 = n - m - 1$ where k is the minimum index of rows that follow a reverse pattern as does row 1. So, by Lemma 5, the number of weak squares encoded in rows that have a reverse pattern as row 1 is

$$\sum_{i=1}^{n-m-1} i = (n-m)(n-m-1)/2.$$

Therefore, the total number of weak squares encoded in W is

$$\sum_{i=1}^{m} i + \sum_{i=1}^{n-m-1} i = (m(m+1) + (n-m)(n-m-1))/2. \qquad \square$$

From Theorem 1, we see that the number of weak squares in a weak square table is quadratic in the string length since m and n are directly proportional to the string length. Then, so is the total number of weak squares in a Sturmian string s, which is the sum of the number of weak squares encoded in $W_o(s)$ and $W_e(s)$.

4. Encodings

In the previous section we showed that the number of weak squares, and hence an algorithm to output the weak squares in a Sturmian string is quadratic in string length. However, with a different output encoding the efficiency of such an algorithm can be improved. In this section we review the existing output encodings for weak repetitions in the literature and show that according to these encodings, an algorithm to compute the weak repetitions in a Sturmian string is quadratic in the string length as well.

Two different encodings are defined in the literature for representing the weak repetitions in a string: C-encoding [5] and R-encoding [6].

4.1. C-Encoding

C-encoding, originally defined by Crochemore [5] to represent the strong repetitions in a string, can be extended to apply for the weak repetitions as well [6]. An entry in C-encoding is a triple (i, p, k) to represent a unique repetition where i is the position of the repetition in the string, and p and k are respectively the period length and order of the repetition. With the addition of a fourth entry to this triple, the repetitions with identical period length and order at consecutive positions can be packed into a single entry. According to this, a 4-tuple (i, p, k, r) represents the repetitions $(i, p, k), (i + 1, p, k), \ldots, (i + r - 1, p, k)$. We call such a 4-tuple a C-*run*. We name the repetition at the leftmost position of the string represented by a C-run as the *base repetition* of that run. Observe that, any C-run can be transformed into exactly one other 4-tuple of which the base repetition is a square.

It is fairly easy to see that the repetitions at odd positions in any Sturmian string s to compose a C-run of which the base repetition is a weak square are represented by the entries $[i, j], [i + 1, j + 1], \ldots, [i + z, j + z]$ of the $W_o(s)$ table where $W_o(s)[i + t, j + t] = 1$ for all $t = 0..z$, and $W_o(s)[i - 1, j - 1] = W_o(s)[i + z + 1, j + z + 1] = 0$.

Consider the words $s_k = (aabaaaba)^k aa$ where $k > 1$. This word is Sturmian, a substring of any Sturmian string X_π where $\pi = \{(3, m), (k', n), \ldots\}$, $m = 3 \pm 1$, $k' = 2k$, $n = k' \pm 1$ [9]. In Table 1, we present $W_o(s_k)$. If we look at $W_o(s_k)$ for a value of k, there are two consecutive 1 and 0 entries one after the other, switching at every two rows. From this pattern, we calculate the number of those sequences of 1 entries in $W_o(s_k)$ which are parallel to the diagonal $i = j$:

$$K = (2k + 1) + 2 \sum_{i=0}^{k-1} (2i + 1) + k = 2k^2 + 3k + 1$$

where $8k + 2 = |s_k|$. Even in the best case when the corresponding strings at even positions of each of these runs are also weak squares so that K is the exact number of C-runs in s, the number of C-runs in s is quadratic in string length.

Table 1
$W_o(s_k)$ table where $s_k = (aabaaaba)^k aa$, $k > 1$

	1 aa	2 ba	3 aa	4 ba	5 aa	6 ba	7 aa	8 ba	9 aa	... ...
1	1	0	0	1	1	0	0	1	1	...
2	—	0	0	1	1	0	0	1	1	...
3	—	—	1	0	0	1	1	0	0	...
4	—	—	—	0	0	1	1	0	0	...
5	—	—	—	—	1	0	0	1	1	...
6	—	—	—	—	—	0	0	1	1	...
7	—	—	—	—	—	—	1	0	0	...
8	—	—	—	—	—	—	—	0	0	...
9	—	—	—	—	—	—	—	—	1	...
...	...	...	...	...	...	...	...	...	...	...

4.2. *R*-Encoding

R-encoding is based on (c, p) pairs. A (c, p) pair represents a weak square of period length p around center c where c is the position in x where the second instance of the period starts. Weak squares centered around the same position with consecutive period length values can be packed into an *R-run* with the extension of (c, p) pairs into (c, p_1, p_2) triples to represent the weak squares $(c, p_1), (c, p_1+1), \ldots, (c, p_2)$ where $p_1 \leqslant p_2$, $x[c - p_i..c + p_i - 1]$ is a weak square for all $p_i \in [p_1, p_2]$.

In W_o table of a Sturmian string, the repetitions at odd positions of the string which are constituents of an R-run are represented by a sequence of consecutive entries in W_o at cells of which the column and row indices add up to the same value, namely the consecutive entries which are parallel to the “reverse” diagonal of the table. Consider again the Sturmian words $s_k = (aabaaaba)^k aa$ where $k > 1$. From the pattern in $W_o(s_k)$ we described earlier, the total number of such sequences of 1 entries in $W_o(s)$ parallel to the “reverse” diagonal $i = j$ is

$$K' = (2k + 1) + 4 \sum_{i=1}^{k-1} i + 2k = 2k^2 + 2k + 1.$$

Again, even in the best case when the corresponding strings at even positions of each of these runs are also weak squares so that K' is the exact number of R-runs in s_k, the number of R-runs in s_k is quadratic in string length since $8k + 2 = |s_k|$. This completes the facts to support the argument that the optimum efficiency of an algorithm to compute the weak repetitions in a Sturmian string s based on an encoding defined in the literature is $O(|s|^2)$.

5. Conclusion

As we noted in Section 3 the set of entire rows of a weak square table of a Sturmian string can be partitioned into two sets so that the elements of one partition are the rows following identical pattern, and the elements of the other are those following a reverse pattern as does row 1. Suppose that $P_o = \{i_1 = 1 < i_2 < \cdots < i_n\}$ is one of these partitions associated to $W_o(s)$ of a finite Sturmian string s. Consider the sets $I_{S_o} = \{i'_k\colon\ i'_k = 2i_k - 1,\ i_k \in P_o\}$ and $J_{S_o} = \{j'_k\colon\ j'_k = 2j_k, \Gamma_{i_1}(W_o(s))[j_k] = 1, j_k \in [1..|\Gamma_{i_1}(W_o(s))|]\}$. I_{S_o} and J_{S_o} are respectively the sets of starting positions and ending positions of all possible weak squares encoded in rows that follow the identical pattern as row 1 in $W_o(s)$. Looking at the tuple $S_o = (i'_k\colon\ i'_k \in I_{S_o}; j'_k\colon\ j'_k \in J_{S_o})$, each (i', j') pair in this tuple where $i' < j'$ represents a weak square of period length $j' - i' + 1$ positioned at i', and the tuple S_o as interpreted for all pairs satisfying this condition specifies the entire weak squares encoded in rows $i_1..i_n$ of $W_o(s)$. Similarly, another tuple S'_o for the other partition of rows in $W_o(s)$ can be constructed so that S_o and S'_o together encode all the weak squares located at odd positions of the string s. With the application of the same technique on $W_e(s)$ table, the weak squares in a finite Sturmian string is encoded in 4 tuples, S_o, S'_o, S_e, S'_e, which can be produced in time linear in string length. An $(i'_1, i'_2, \ldots, i'_m; j'_1, j'_2, \ldots, j'_n)$ tuple can further be reduced in time linear in the number of weak squares in the string to m distinct tuples: $(i'_k; j'_k, j'_{k+1}, \ldots, j'_n)$ each of which representing the entire weak squares located at position i'_k of the string.

In this study, we analyzed the Sturmian strings based on their balance property. Future studies may include the consideration of additional combinatorial properties of Sturmian strings to find out other encodings to improve the efficiency of an algorithm to compute the weak repetitions in these strings. Another line of study may be conducted to investigate the representations of Sturmian strings from the perspective we presented in this paper to connect these representations to our analysis.

Acknowledgements

I wish to thank Jeffrey Shallit for his review in the preliminary stages of this study.

References

[1] J.-P. Allouche, J.O. Shallit, Automatic Sequences: Theory, Applications, Generalizations, Cambridge Univ. Press, Cambridge, to appear.

[2] J. Berstel, Recent Results in Sturmian Words, Developments in Language Theory, World Scientific, Singapore, 1996, pp. 13–24.

[3] T.C. Brown, Descriptions of the characteristic sequence of an irrational, Canad. Math. Bull. 36 (1993) 15–21.

[4] E.M. Coven, G.A. Hedlund, Sequences with minimal block growth, Math. Systems Theory 7 (2) (1973) 138–153.

[5] M. Crochemore, An optimal algorithm for computing the repetitions in a word, Inform. Process. Lett. 12 (5) (1981) 244–250.

[6] L.J. Cummings, W.F. Smyth, Weak repetitions in strings, J. Combin. Math. Combin. Comput. 24 (1997) 33–48.

[7] A. De Luca, Sturmian words: string, combinatorics, and their arithmetics, Theoret. Comput. Sci. 183 (1997) 45–82.

[8] A. De Luca, F. Mignosi, Some combinatorial properties of Sturmian words, Theoret. Comput. Sci. 136 (1994) 361–385.
[9] A. Karaman, Representations of Sturmian strings, Master of Science Thesis, Department of Computing and Software, McMaster University, 1998.
[10] A. Karaman, Weak repetitions in Sturmian strings, Technical Report: CAS99-02, Department of Computing and Software, McMaster University, 1998.
[11] F. Mignosi, On the number of factors of Sturmian words, Theoret. Comput. Sci. 82 (1991) 71–84.